

APPENDIX K

Input: 1. A dataset consist of continuous and dummy variables, it is normalized, X

2. Target variable, y

Output: X, some variable might be dropped in the process

Parameter : Threshold of correlation, TC. Default 0.8. Range : 0.8~0.95.

Process:

NE = $X^T \cdot X$; //NE is the normal equation matrix, each element is in its absolute //value
While there exists any element $\text{abs}(\text{NE}(i,j)) > \text{TC}$

cor1 = absolute value of correlation between x_i and y;

cor2 = absolute value of correlation between x_j and y;

If $\text{cor1} > \text{cor2}$

 Mark x_j as dropped

 Fill 0s in j th row and j th column of NE;

Else

 Mark x_i as dropped

 Fill 0s in i th row and i th column of NE;

End If

End While

Delete variables in X that are marked to be dropped.

Delete the corresponding rows and columns in the normal equation matrix NE.

Store names of the dropped continuous and dummy variables